DOCUMENT RESUME

ED 365 692 TM 020 860

AUTHOR Pigge, Fred L.; Marso, Ronald N.
TITLE A Summary of Published Research: Classroom Teachers' and Educators' Attitudes toward and Support of Teacher-Made Testing. [Revised.]
PUB DATE Oct 93
NOTE 29p.; Paper presented at the Annual Meeting of the Midwestern Educational Research Association (Chicago, IL, October 13-16, 1993). For a related paper, see ED 346 148.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classroom Techniques; Educational Attitudes; Educational Practices; Educational Research; *Elementary School Teachers; Elementary Secondary Education; Literature Reviews; Research Reports; *Secondary School Teachers; Student Evaluation; *Teacher Attitudes; Teacher Education; *Teacher Made Tests; Teaching Methods; Test Construction; *Test Use

ABSTRACT

This paper presents a summary of findings from a review of approximately 225 studies of K-12 classroom teachers' attitudes toward and other educators' attitudes and support of teacher-made tests and testing practices. The findings from the review indicate classroom teachers have a positive attitude toward teacher-made tests and regard these tests as having a far more positive impact upon their day-to-day instruction than do other types of tests. Further, teachers' positive regard for these tests is reflected in their heavy reliance upon and frequent use of these self-constructed tests in their classrooms. In contrast, other educators express a positive attitude toward teacher-made tests and testing in K-12 classrooms, but this attitude is not reflected in the limited extent to which preservice and inservice training and other basic resources such as test typing, duplication, and scoring services are made available to teachers in meeting their day-to-day testing responsibilities. Four tables are included. (Contains 71 references.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
FRED L. PIGGE
RONALD N. MARSO
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 365 692

A Summary of Published Research: Classroom Teachers' and Educators' Attitudes Toward and Support of Teacher-Made Testing

Fred L. Pigge and Ronald N. Marso
College of Education and Allied Professions
Bowling Green State University
Bowling Green, Ohio 43403

A paper presented at the Annual Meeting of the
Midwestern Educational Research Association

Chicago
October 13-16, 1993

Abstract

This paper presents a summary of findings from a review of approximately 225 studies of K-12 classroom teachers' attitudes toward and other educators' attitudes and support of teacher-made tests and testing practices. The findings from the review indicate classroom teachers have a positive attitude toward teacher-made tests and regard these tests as having a far more positive impact upon their day-to-day instruction than do other types of tests. Further, teachers' positive regard for these tests is reflected in their heavy reliance upon and frequent use of these self-constructed tests in their classrooms. In contrast, other educators express a positive attitude toward teacher-made tests and testing in K-12 classrooms, but this attitude is not reflected in the limited extent to which preservice and inservice training and other basic resources such as test typing, duplication, and scoring services are made available to teachers in meeting their day-to-day testing responsibilities.

TM020860

BEST COPY AVAILABLE

A Summary of Published Research: Classroom Teachers' and Educators' Attitudes Toward and Support of Teacher-Made Testing

Most K-12 teachers feel that teacher initiated assessments have a major impact upon pupil learning and spend considerable time in various forms of evaluation each class day. Teachers believe that classroom tests guide and instigate pupil learning efforts (Rogers, 1969), that the nature of classroom tests influences their pupils' study habits (D'Ydewalle, Swerts, & Decorte, 1983), that testing frequency influences pupil achievement (Baugert-Downs, Kulik, & Kulik, 1988), that carefully administered, announced, and monitored classroom tests produce higher pupil performance (Hill & Wigfield, 1984), and that prompt return of classroom tests accompanied by the provision of knowledge of results increases pupil achievement (Kulik & Kulik, 1988).

Although classroom teachers frequently use and strongly believe in the positive benefits of teacher-made tests and educators believe testing and evaluation is one of the most potent forces influencing education, Crooks (1988) contends that the actual elements of the evaluation process in the K-12 classroom have received less attention from researchers than have many other aspects of education including standardized testing. Similarly, Stiggins, Conklin, and Bridgeford (1986) described the existing research related to classroom tests and testing practices to be limited and narrow in scope.

The purpose of this paper is to provide a bibliography and selected findings from a more extensive review of the research literature addressing K-12 classroom teachers' skills and knowledge related to the development and use of teacher-made tests. The full report of the findings from this review is scheduled to appear as a chapter in Teacher Training in Assessment, Steven Wise editor, in volume seven of the Buros Nebraska Symposium in Measurement and Testing. The present paper provides information related to just the following two of the several questions addressed in the more extensive literature review: 1) What attitudes do educators have toward teacher-made tests and testing practices as revealed through various self report procedures and as revealed through the extent of training, support, and resources made available for these activities? 2) What attitudes do classroom teachers have toward teacher-made tests and testing practices as revealed through various self report procedures, their testing practices, and analyses of their self-constructed tests?

The research studies reviewed for the larger study were identified through computer searches of the ERIC data base and through the gathering of those reports cited within the computer-identified studies. These procedures resulted in the collection of approximately 225 research reports.

Question One:

Educators' Attitudes Revealed Through Self Reports and Availability of Training and Resources for Classroom Testing

Testing Standards and Codes

Until the standards for K-12 classroom teacher competence in the assessment of pupils (NCME-ASCTE-AFT-NEA) were published in 1990, the testing community had not provided clear expectations of or standards for classroom teachers' testing competence. In contrast, the statements of standards for standardized testing can be traced back to the mid-twentieth century and are currently conveyed in the Standards for Educational and Psychological Testing which were jointly developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (AERA-APA-NCME, 1985). More recently these latter standards were supplemented by the 1988 Code of Fair Testing Practices in Education also jointly sponsored by these three professional associations. The Code was designed to complement the earlier standards and differs from the standards in audience addressed and purpose. It is focused just upon standardized educational testing but addresses the practices of both test developers and test users. Its stated primary role is to address test and test score misuses which have tended to generate far more public criticism than have questions about test quality itself (Diamond & Fremer, 1989).

Neither the 1988 Code nor the 1985 Standards address teacher-devised testing. Frisbie and Friedman (1987) did make an effort to show a relationship between the 1985 Standards and teacher-devised testing; however, the results of their effort were illustrative rather than enumerative in scope. Thus, it appears that the measurement community has provided less professional guidance for and, as noted previously, less research of teacher-made testing than it has for standardized testing. This relative neglect of teacher-devised testing has occurred in spite of the fact that the measurement profession perceives teacher-made tests and not standardized tests to be the dominant influence in K-12 classrooms (Stiggins, 1985).

Even though the measurement community appears to have provided less research support and professional guidance for teacher-devised testing in contrast to standardized testing, it appears to have considerable doubts about the testing knowledge, skills, and practices of educators. For example, Diamond and Fremer (1989) noted that the Institute for Research on Teaching, which coordinated the development of the previously described fair testing code, was particularly critical of the inadequate training of educational personnel relative to the interpretation and use of tests.

Testing Resources

The perceptions of the extent to which testing expertise and other resources are available to support teacher-devised testing

activities in the K-12 schools appears to be as bleak as the measurement community's perceptions of the adequacy of teachers' testing competencies. Ruddell (1985), after conducting interviews of school principals and school district central office staff relative to the availability of testing expertise in K-12 schools, concluded that they possessed very limited knowledge about tests and test score interpretation concepts such as the standard error of measurement.

Marso and Pigge (1990) conducted a survey of school district designated directors of standardized testing and found that many school testing directors, themselves, have limited training in testing and evaluation. Further, many of the testing directors when queried about support services which they provided for classroom teachers, contrary to the expectations stated in the Standards for Educational and Psychological Testing, reported that they were not responsible for encouraging the use of standardized test results in their schools, for training teachers to proctor standardized tests, and for training teachers to better interpret scores from standardized tests.

Marso and Pigge also found that many of the testing directors reported increased demands on their time resulting from added responsibilities for the management of mandated statewide pupil competency testing; thus, undoubtedly, also reducing the testing directors' opportunities for providing teachers with testing expertise or related testing support services. These researchers concluded that it is probably safe to assume that if testing directors do not provide basic testing support services for teachers, then these essential services probably are not being provided in the schools. This conclusion was based partly on the assumption that no one else in these schools, especially in smaller size school districts, would likely have this responsibility or would likely have the expertise to deliver such services.

Relatedly, Stiggins (1985) noted that few school administrators have the training or the experience necessary to help teachers with classroom testing or related responsibilities. As further evidence of this lack of expertise, Marso and Pigge (1989c) reported negative correlations between principals' and supervisors' ratings of teachers' various question type writing skills (e.g., ability to write multiple-choice and other types of questions) and the observed levels of the adequacy of teachers' various question writing skills as displayed on their self-constructed tests. As the adequacy of the teachers' question writing skills in this study was judged upon the frequency of violations of common test construction guidelines, this finding may suggest that school administrators, who themselves tend to have little or no training in testing, may not have sufficient awareness of common test item flaws to be able to identify question writing violations in teacher-constructed tests let alone effectively advise teachers how to avoid these violations.

Lambert (1980-81) collected opinions about teachers' attitudes, training, and knowledge about teacher-made and standardized tests from a national sample of state legislators, state teacher association officials, and deans of colleges of education. He found both

agreement and divergence between and within these three samples. For example, approximately one-third of the deans reported that their colleges did not offer a measurement course for their teacher candidates and that they had no intention of doing so; nevertheless, the deans agreed with one another that classroom teachers have a negative attitude toward standardized tests, that teachers should know more about tests, and that it is very important for teachers to construct superior tests for the assessment of their pupils. Ultimately, Lambert concluded that all three groups sampled needed to know more about the value and limitations of tests.

In regard to the extent to which resources are made available to support teachers' testing activities, Marso and Pigge (1988c) asked over 800 K-12 teachers, principals, and supervisors to report the extent to which selected resources were available in their schools to support classroom teachers' testing responsibilities. They found that even basic typing and duplication services were not consistently available in 50% of the schools, grade assignment guidelines were not available in 50% of the schools, and basic computer services (e.g., test scoring, item pools, item analyses, etc.) were not available in approximately 75% of the schools.

Dorr-Bremme (1983), after using questionnaire and interview procedures to gather data from a national sample of school staff in 114 school districts, reported that most K-12 teachers do not receive inservice training or assistance of other types in selecting, developing, and use of tests. Rather significantly, this researcher found a relationship between teachers' attitude toward school testing and the extent to which school support for testing was made available in forms such as expressed principal interest, test interpretation assistance, and inservice training related to testing. In school districts where these testing support services were more extensive, teachers' attitude toward testing was positive; in school districts where these resources and services were very limited, teachers' attitude toward testing was less positive. In another study related to the availability of support for testing, Gullickson (1984) also found that teachers reported having little assistance available for the preparation, analysis, scoring, or interpretation of teacher-made tests.

Training Resources

Hermanowicz (1980) argued that a major component in K-12 teachers' preservice education ought to be training in the development and use of classroom tests. Practicing teachers, themselves, report that assessment of pupils is a key element in the instructional process, and measurement specialists such as Stiggins, Conklin, and Bridgeford (1986) and Dorr-Bremme (1983) have provided information describing how classroom teachers do integrate testing within their day-to-day instructional practices. Further, Schafer and Lissitz (1987) reported an increasing awareness of the importance of teachers' pupil assessment skills within the educational community as evidenced by the positive positions taken by the two major national teacher organizations on pupil assessments and by the inclusion of testing as

one of the five skill components measured by the recently revised National Teachers Examination.

Despite this evidence of the educational community's awareness of teacher need for pupil assessment competencies, considerable evidence exists which indicates that a significant proportion of professional school personnel receive little or no formal training in measurement and evaluation. After conducting a survey of 438 institutions of higher education, Schafer and Lissitz (1987) found that approximately only one-third of various K-12 educational personnel preparation programs required a measurement course for certification. Even more disconcerting, they found that approximately just 25% of the elementary and secondary teacher preparation programs required a measurement course. They further noted that, although administrators are expected to serve as instructional leaders in schools, the administrator education programs were least likely of all preparation programs to require measurement training. Among the advanced certification programs for educators, they found that only the counseling programs are very likely to have a measurement course requirement.

Gullickson and Hopkins (1987) conducted a regional survey of 99 colleges of education and found that approximately one-half of the colleges provided a measurement course for their preservice teachers; whereas the other colleges provided just a unit of instruction in measurement within another course. In an earlier study Roeder (1973) surveyed 860 colleges of education and found that somewhat fewer than one-half of their elementary teacher preparation programs required a separate tests and measurement course.

Relatedly, Green and Williams (1989) found that classroom teachers with more training in measurement reported scheduling teacher-made tests more frequently and using the results of standardized tests more extensively than did teachers with less training. A rather disturbing finding by these researchers was that the less well trained teachers perceived themselves to be more knowledgeable about interpreting the results of tests than did the better trained teachers. In a similar earlier study, Green and Stager (1986-87) found that the extent of teachers' training in testing did not influence the frequency of their use of teacher-made tests; however, they did find that the more well trained as compared to the less well trained teachers were more likely to use appropriate test development practices such as item analysis and test specification tables.

Educators typically avoid measurement training when not required in their preparation programs (Coffman, 1983; Schafer & Lissitz, 1987; Stiggins & Bridgeford, 1982). Some individuals have suggested that educators may avoid measurement training because the training being provided has not been designed to meet practical classroom needs Airasian & Madaus, 1983; Stiggins & Bridgeford, 1985). In support of this speculation, Gullickson (1986a) found discrepancies between common college measurement course topics and practicing teachers' perceptions of what testing topics and skills are needed to

successfully function in their classrooms. He reported that classroom teachers rely heavily upon informal observations of and direct communications with pupils in making instructional decisions, and they perceive little need for statistical testing procedures. In contrast, Gullickson noted that preservice measurement instruction tends to focus upon paper and pencil measurement assessments and statistical analyses of data rather than upon informal data gathering procedures.

The findings from several other studies also indicate that discrepancies do exist between K-12 classroom teachers' testing practices and typical educational measurement training. For example, Gullickson and Ellwein (1985) and Marso and Pigge (1991) found that few practicing teachers use statistical analysis procedures in interpreting pupil test performance. Kellaghan, Madaus, and Airasian (1982) reported that measurement training has resulted in little real impact upon teachers' testing practices and suggested that it is unlikely to do so until it focuses on the actual demands of pupil assessment in K-12 classrooms. Further complicating these concerns about measurement training for classroom teachers, Gullickson and Hopkins (1987) found that many preservice measurement professors, themselves, have limited measurement training and/or experience in the use of tests in K-12 classroom settings.

In addition to the major concerns about K-12 teachers having little or no preservice training in testing and whether such training is appropriate, several researchers have reported that inservice teacher training in testing is almost nonexistent (Dorr-Bremme, 1983; Gullickson, 1984), and Marso and Pigge (1991) found indirect evidence of this in that neither teachers' ratings of their own testing proficiencies nor the observed quality of their teacher-made tests differed when these ratings and tests were grouped by the teachers' years of teaching experience. Further, teachers frequently perceive their inservice training to be not very helpful. For example, Marso and Pigge (1987b) found that of all school experience factors assessed, first-year teachers were most disappointed with their inservice training. Relatedly, Stiggins (1988) suggests that teachers will seek inservice training designed to improve their tests and testing practices but will avoid inservice measurement training if it is perceived to be like that provided in preservice training.

In conclusion and as summarized in Table 1, it is apparent that K-12 teachers are perceived by the educational and measurement communities to have limited testing knowledge and skills; that neither measurement consultative expertise nor inservice training in testing is generally available to teachers in most schools; that even basic testing support services such as typing and duplication assistance are not commonly available to teachers in a large number of schools; that a large portion of classroom teachers have had little or no formal preservice or inservice measurement training; and that much of the training in pupil assessment which is available to teachers and teacher candidates is perceived by practicing teachers to be inappropriate relative to actual classroom instructional needs.

- - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - -

Question Two:

Teachers' Attitudes Revealed Through Their Self Reports, Testing Practices, and Self-Constructed Tests

Testing Beliefs and Practices

Mehrens and Lehmann (1987) have estimated that a typical pupil will take between 400 and 1000 teacher-made tests during their K-12 school years, Crooks (1988) and Haertel (1986), have indicated that approximately 5 to 15 percent of a typical classroom day is devoted to pupil assessment, Newman and Stallings (1982) and Stiggins (1988) have estimated that teachers spend approximately 11 to 20 percent of a typical work day in some aspect of pupil assessment, and Marso and Pigge (1991) reported that K-12 teachers construct an average of 54.6 formal paper and pencil tests in a typical school year.

Teachers use both self-constructed tests and publisher-constructed (textbook or workbook) tests but prefer their own tests. Dorr-Bremme (1983), studying a national sample of teachers, reported that 95 percent of the teachers used self-constructed tests and 77 percent used publisher-constructed tests. But regardless of test source, teachers and pupils spend considerable classroom time and effort in testing activities (Fleming & Chambers, 1983).

Teachers' opinions about what are appropriate testing practices appear to vary somewhat by grade level of instruction and by subject area content being assessed. At the upper grade levels, teachers rely more on teacher-constructed as compared to publisher-constructed tests, express more concerns about the quality of pupil assessments, and use somewhat more test quality control procedures such as item analysis and checks on reliability than do teachers in the lower grades (Marso & Pigge, 1991; Stiggins & Bridgeford, 1985). Primary grade teachers place more focus on pupil work samples than upon testing; lower elementary grade teachers more frequently use worksheets and tests provided in publisher textbooks and workbooks than do teachers at other grade levels; and upper grade and high school teachers predominantly use formal self-constructed tests in their assessment of pupils (Herman & Dorr-Bremme, 1982; Marso & Pigge, 1991; Salmon-Cox, 1981).

Essay questions and tests appear not to be held in high regard by teachers and are very seldom used at any grade level. Essay questions are more frequently found in English, history, and social studies tests than in other subject area tests; and they are more frequently used in the upper grades than in the lower grades. Math and science teachers test their pupils more frequently but are less likely to use essay items than other subject area teachers. Math and science teachers are more likely to use formal paper and pencil tests than informal assessments. Teachers in writing and speech classes are more

likely than are other teachers to use direct observations and informal judgments in assessing the progress of their pupils (Marso & Pigge, 1991; Stiggins & Bridgeford, 1985).

Upper grade level teachers believe that letter grades or marks should be based primarily on pupil test performance and daily work; whereas K-4 grade teachers believe that daily work and observations are more important than tests in assigning grades. Most teachers when assigning marks consider teacher-devised tests to be a primary source of information (Marso, 1986; Shulman, 1980).

Teachers generally favor self-constructed items as compared to items from other sources, and they typically report constructing from 50 to 75 percent of the test questions used on their tests. Teachers also favor the use of a variety of test items with an average of 2.6 question types found on a typical teacher-devised test (Dorr-Bremme, 1983; Marso & Pigge, 1991; Yeh, 1981).

A combination of completion or short-response type questions followed by matching, multiple-choice, true-false, and essay type questions are most frequently used in teacher-devised tests. When teachers are asked to rate the usefulness, adaptability, and fairness to pupils of the various question types, the question types are placed in a somewhat different order: matching, completion, short-response, multiple-choice, true-false and essay. Teachers do believe that pupils study more for essay tests as compared to objective tests and that essay tests are more likely to function at higher cognitive levels than are objective tests even though they deem the essay items to be less useful and seldom use them (Coffman, 1971; Marso, 1985).

Most classroom teachers provide information to pupils regarding their performance following the administration of a classroom test, and typically they report spending about one-half of a class period for that purpose. Teachers report that pupils usually are very attentive and motivated during these test feedback sessions (Haertel, 1986). Teachers tend to reuse their tests without analysis and revision and seldom use statistical procedures to assess the quality of their tests (Gullickson & Ellwein, 1985; Marso & Pigge, 1988b).

Few studies describe exactly how teachers use tests in their classroom instruction (Kuhs et al., 1985). Teachers, themselves, report a heavy reliance on teacher-made tests in their day-to-day instruction. In contrast they report placing little reliance on standardized tests for making instructional decisions (Salmon-Cox, 1981). Relatedly, Borg, Worthen and Valcarce (1986) reported unfavorable and indifferent teacher attitudes toward the classroom use of standardized tests but highly positive teacher attitudes toward the use of teacher-made tests. Stiggins and Bridgeford (1985) reported that classroom teachers did use their self-constructed tests for pupil diagnosis, grouping, grading, evaluation, and the reporting of pupil progress in their classrooms and that they placed more emphasis upon structured performance assessments than upon spontaneous observations of pupils in making instructional decisions.

A common criticism of teachers is that they tend to over value test scores, and in particular standardized test scores, relative to other available information about pupils. Hall, Carroll, and Comer (1988) found, however, that classroom teachers consistently favored the results of their self-constructed tests over the results of standardized or state competency tests in making decisions. They also noted that teachers made decisions with a reasonable regard for the complex data requirements present in classroom settings. Similarly, Lazar-Morrison, Polin, Moy, and Burry (1980) concluded that teachers place greater confidence in the results of their own judgments of pupil performance than upon any formal tests; and Stiggins and Bridgeford (1985) reported that teachers rely on a number of sources of information in making decisions about pupils and that teachers' relative reliance on sources of pupil information is in the following order: teacher-made tests, standardized tests, structured performance assessments, and spontaneous observations.

Dorr-Bremme (1983) concluded that teachers bring several types of assessments to their decisions about pupils and that they rely more on personal experiences and observations than upon test scores. Similarly, Salmon-Cox (1981) reported that high school teachers made very little use of standardized test scores in evaluating pupils; Shavelson, Cadwell and Izu (1977) found that teachers gave due consideration to the reliability of data in making decisions about pupils; and Kellaghan, Madaus, and Airasian (1982) found that teachers can accurately predict pupil test performance and only use students' standardized test scores to corroborate their own judgments.

The findings of the research related to teachers' use of test scores suggests that classroom teachers use scores to raise but not to lower their expectations of individual pupils. When teachers note a discrepancy between their perceptions of a pupil's ability and test scores, teachers ignore test scores when the scores suggest that less might be expected of a pupil, and teachers raise their expectations of a pupil when test scores suggest that more might be expected of a pupil (Airasian, Kellaghan, Madaus, & Pedulla, 1977).

Two studies of teachers' attitudes toward educational testing appear to be representative of teacher perceptions of tests and testing. Green and Stager (1986-87) surveyed 555 classroom teachers and reported that younger teachers are more skeptical of testing than older teachers, that upper grade teachers are more positive toward testing than lower grade teachers who place more emphasis on classroom observations and informal pupil assessments, that teachers are positive toward teacher-made tests but tend to be negative or indifferent about standardized tests, that most teachers express interest in upgrading their testing skills, and that reported use of contemporary measurement practices (e.g., use of test specification tables and item analysis, etc.) was found to be somewhat related to more frequent pupil testing practices but not to attitude toward testing.

In a second study of teachers' attitudes and beliefs about tests, Gullickson (1984) reported that teachers felt that teacher-constructed

tests result in increased pupil effort, influence pupil self-concept, create desirable competition among students, improve interaction among pupils, improve the classroom learning environment, better focus teaching, provide a good learning experience for pupils, motivate pupil study, and accurately reveal pupil progress. Further, Gullickson found that teachers believe that: frequent brief tests are more desirable than infrequent lengthy tests, school administrators encourage frequent testing of pupils, pupils prefer frequent tests, pupils try hard on tests, tests are an important instructional tool, tests need to be tied closely to instruction, tests help evaluate instruction, essay tests better assess pupil progress than objective items and measure at a higher cognitive levels, tests should not be the sole determinant of grades, and that tests are necessary to help justify grades to parents.

Pupils appear to reflect the attitudes of their teachers about tests, for students also feel that tests help them learn, and they too favor frequent testing. Pupils report that teacher-made tests must be taken more seriously and are more difficult than standardized tests (Kulik & Kulik, 1981), and, like many teachers, some pupils feel that standardized tests are a waste of time (Stetz & Beck, 1981).

In summation, this review of teachers' testing practices and beliefs suggests that K-12 classroom teachers appear to have a very favorable attitude toward teacher-made tests and testing: they feel it is appropriate to and expend considerable effort and time in fulfilling teacher instigated testing responsibilities in their classrooms; they favor and schedule tests frequently followed by class discussions of pupil performance; they do have concerns about but also positive feelings about the role of testing and pupil evaluation in the instructional process; and they have confidence in their classroom tests and their overall testing ability but recognize that they would benefit from practical training in testing. A summary of teachers' testing practices, beliefs, and attitudes is presented in Table 2.

- - - - - - - - - - - - - - -
Insert Table 2 about here
- - - - - - - - - - - - - - -

Assessments of Teacher-Made Tests

Few studies of teachers' testing knowledge and skills have been conducted wherein direct analyses of samples of their teacher-made tests have served as the major data gathering procedure. One such study was reported by Fleming and Chambers (1983). They analyzed 342 teacher-made tests encompassing 8,800 test questions constructed by teachers assigned to several grade levels and subject areas in the Cleveland Public Schools. Some of the more salient findings from this study follow:

1. Short-answer (including fill-in-the-blank) questions were most frequently used followed by matching, multiple-choice, true-false (seldom used), and essay questions. Essay items were very infrequently found on any of these teachers' tests (about 1% of all questions).

2. Almost 80 percent of the questions found on the tests measured at the knowledge level. The higher level functioning items, however, rather than being spread equally throughout all the tests, were found primarily on the math tests. Few questions on any tests were judged to measure pupils' ability to make applications.

3. Fewer than two-thirds of the tests contained directions for all question types, and even though questions were grouped by question type on all tests, they often were not numbered consecutively or were not numbered at all.

4. Indicative of inadequate support services, many of the tests were handwritten, were poorly reproduced, and had pages over-crowded with content.

In a second study, Marso and Pigge (1991) analyzed 6504 test questions contained within 455 question exercises (a group of questions of similar type on a test) found on 175 formal teacher-made tests constructed by classroom teachers with from one to 10 years of teaching experience who had completed a preservice tests and measurement course. Some of the more salient findings from this study follow:

1. Question type use varied by grade level and subject area content with an "average" test made up of 2.6 item types. The question types used from highest to lowest frequency were short-response, matching, true-false, multiple-choice, problems, completion, interpretive exercises, and essay.

2. Very few differences were noted in test construction practices or test construction quality when the tests were examined in terms of whether the teachers constructing the tests had few or more years of teaching experience.

3. Matching exercises were found to be the most error prone question type.

4. Teachers reported preparing an average of 54.6 formal teacher-made tests each year, approximately 70 percent of the teachers scheduled a test once every two weeks or more frequently in a typical class, and over 50 percent of the teachers reported writing three-fourths or more of the questions used on their tests.

5. As a total group of questions on all tests, 72 percent were judged to be functioning at the knowledge cognitive level, but the large majority of the questions functioning beyond the knowledge level were contained just in the math and science tests.

In a study of secondary math and science teacher-constructed tests, Oescher and Kirby (1990) analyzed 34 tests containing over 1400 test questions and gathered the responses of 35 teachers to a teacher

testing practices questionnaire. They concluded that 70 percent of the tests contained format errors, 26 percent of the tests were without directions, and all question types present on the tests were judged to violate several basic item writing guidelines.

In other studies of less comprehensive samples of teacher-made tests, Billeh (1974) analyzed 33 science tests to determine their cognitive functioning levels and reported that of all questions reviewed 72 percent functioned at the knowledge level, 21 percent functioned at the comprehension level, and seven percent functioned at the application level. The more experienced teachers in Billeh's sample used more knowledge level items, but no differences in the cognitive functioning levels of the tests were found when classified by grade level or by extent of teacher training. Black (1980) reported an analysis of 48 secondary level science tests and found that the cognitive functioning levels of the tests varied within the science subject areas. Biology tests contained 94 percent knowledge, chemistry 66 percent knowledge, and physics 56 percent knowledge level questions.

Ball, Doss, and Dewalt (1986) studied the tests constructed by 74 junior and senior high social studies teachers. They found that, although approximately 75 percent of these teachers indicated that higher level instructional objectives were most important to student learning and approximately 25 percent of these teachers reported that they predominantly used these higher level type objectives in their teaching, 98 percent of the questions on these teacher-made social studies tests were measuring only at the recall level. Marso and Pigge (1991) also found that the social studies tests collected in their study were composed of questions measuring almost exclusively 100 percent at the knowledge level.

Similarly, Stiggins, Griswold, and Wikeland (1989) conducted interviews, class observations, and direct analyses of teacher-constructed tests of 36 K-12 classroom teachers who had been participating in inservice teacher training focused on school district endorsed efforts to teach with a focus on the development of their pupils' thinking skills. They found that all of these teachers' self-constructed tests were composed of questions functioning 100 percent at the knowledge level except for the math tests. These researchers commented that it was easier to train teachers to teach with a focus on their pupils' higher thinking levels than it was to train teachers to design tests to measure pupil achievement at these higher levels.

In summation, the review of studies of the direct analyses of teacher-constructed tests have provided further suggestions about teachers' attitudes about classroom tests and testing practices. Despite teachers' favorable attitude towards teacher-made tests, their tests commonly are lacking in general and technical qualities. Teachers appear not to have time/resources, appear not to have sufficient knowledge or skills, or appear not to sufficiently value format and item writing guidelines and test quality check/improvement practices. Teacher-made tests reveal frequent violations of the most

commonly accepted question and test format writing guidelines, and teachers' self-constructed tests appear not to improve with increasing years of teachers' teaching experience. A summary of the more specific suggestions about teachers' attitude toward testing revealed through analyses of their teacher-made tests are presented on Table 3.

- - - - - - - - - - - - - -
Insert Table 3 about here
- - - - - - - - - - - - - -

Teachers' Beliefs about Testing Skills and Training

Teachers report that they place more reliance on informal in contrast to formal assessments in making K-12 classroom decisions (Gullickson, 1985; Linn, 1990; Salmon-Cox, 1981). Teachers indicate that they need to improve their test construction skills, but they report little need for measurement statistics or for knowledge of legal issues associated with testing in K-12 classrooms (Gullickson, 1986a). Teachers perceive both teacher-made tests and informal observations of pupils to be useful in making day-to-day instructional decisions, but they consider previous teaching experiences to be more useful than test scores in planning instruction for the school year (Dorr-Bremme, 1983).

Borg, Worthen, and Valcarce (1986) and Marso and Pigge (1987a) found that K-12 classroom teachers rated more highly their need for measurement skills closely associated with instruction than their need for skills such as writing structurally sound test questions. Similarly, Newman and Stallings (1982) found that teachers reported heavy reliance upon their self-constructed tests for making decisions about activities most closely related to instruction such as diagnosing pupil strengths and weaknesses, assessing pupil progress, and assessing pupil mastery of units of instruction.

The data presented in Table 4 are illustrative of classroom teachers' ratings of their need for a variety of testing competencies (Marso & Pigge, 1987a). As did the teachers in previously noted studies, these classroom teachers reported relatively little need for measurement statistics. The teachers reported a high need for competencies involving instructional use of test results and for those competencies related to assurances of test validity. They also reported a rather low need for question writing skills which could be deemed necessary to attain the test validity and test instructional uses which they rated highly. Similarly, the teachers rated rather low the need for competency in selecting good test questions from sources such as teacher manuals.

- - - - - - - - - - - - - -
Insert Table 4 about here
- - - - - - - - - - - - - -

Although there is some inconsistency in the research findings about teachers' perceptions of their own testing ability, teachers typically: rate the effectiveness of their training in testing

somewhat below the training they received in other professional areas (Gullickson, 1984; Marso & Pigge, 1987a), rate their testing proficiencies somewhat lower than their proficiencies in other professional knowledge or skill areas (Marso & Pigge, 1987a), and express concern about the level of their testing skills and believe that they could benefit from practical training in tests and measurements skills (Crooks, 1988; Haertel, 1986). Relatedly, first-year teachers rank the extent of their concerns about pupil evaluation and assessment above all other professional concerns except for their concerns about classroom management, pupil motivation, and coping with individual differences among pupils (Veenman, 1984).

Teachers commonly do not feel confident about their ability to write good test questions (Carter, 1984; Gullickson, 1985; Stiggins & Bridgeford, 1985) and are uncertain about how to improve their tests (Carter, 1984). Teachers report that they believe many of their questions and concerns about testing could be alleviated through training (Carter, 1986). Conversely, several researchers have reported that teachers express confidence in their tests as well as in their overall testing knowledge and do not want more training in testing (Green & Stager, 1986-87).

This apparent conflict in findings, which suggests that teachers seemingly both desire but do not want more training in testing, may have been explained at least in part by Stiggins (1988). He noted that teachers do often express confidence in their tests and in their general testing knowledge, but they are uncertain about the technical aspects of testing and they do want practical help in improving their tests and their testing practices. What teachers do not want, he concluded, is more of the theoretical-impractical training typically associated with university tests and measurement courses and workshops.

PC:13A

BEST COPY AVAILABLE

References

AERA-APA-NCME. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Airasian, P. W., Kellaghan, T., Madaus, G. F., & Pedulla, J. J. (1977). Proportion and direction of teacher rating changes of pupil progress attributable to standardized test information. Journal of Educational Psychology, 69(6), 702-709.

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. Journal of Educational Measurement, 20(2), 103-119.

Ball, D. W., Doss, A. R., & Dewalt, M. W. (1986). Level of teacher objectives and their classroom tests: Match or mismatch. Journal of Social Studies Research, 10(2), 27-31.

Bangert-Drowns, R., Kulik, J., & Kulik, C. (1988). Effects of Frequent Classroom Testing. Unpublished manuscript, University of Michigan.

Billeh, V. Y. (1974). An analysis of teacher-made test items in light of the taxonomic objectives of education. Science Education, 58, 313-319.

Black, T. R. (1980). An analysis of levels of thinking in Nigerian science teachers' examinations. Journal of Research in Science Teaching, 17, 301-306.

Borg, W. R., Worthen, B. R., & Valcarce, R. W. (1986). Teachers' perceptions of the importance of educational measurement. Journal of Experimental Education, 55(1), 9-14.

Carter, K. (1984). Do teachers understand the principles for writing tests? Journal of Teacher Education, 35, 57-60.

Carter, K. (1986). Test-wiseness for teachers and students. Educational Measurement: Issues and Practice, 5, 20-23.

Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed., pp. 271-302). Washington, DC: American Council on Education.

Coffman, W. E. (1983). Testing in the schools: A historical perspective. Paper presented at the annual invitational conference of the Center for the Study of Evaluation, University of California, Los Angeles.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58, 438-481.

Diamond, E. E., & Fremer, J. (1989). The joint committee on testing practices and the code of fair testing practices in education. Educational Measurement: Issues and Practices, 8, 23-27.

Dorr-Bremme, D. W. (1983). Assessing students: Teachers' routine practices and reasoning. Evaluation Comment, 6, 1-12.

D'Ydewalle, G., Swerts, A., & DeCorte, E. (1983). Study time and test performance as a function of test expectations. Contemporary Educational Psychology, 8, 55-67.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E Hathaway (Ed.), Testing in the Schools: New Directions for Testing and Measurement, (No. 19, pp. 29-38). San Francisco: Jossey-Bass.

Frisbie, D. A., & Friedman, S. J. (1987). Test standards: Some implications for the measurement curriculum. Educational Measurement: Issues and Practice, 6(3), 17-23.

Green, K. E., & Stager, S. F. (1986-87). Testing: Coursework, attitudes, and practices. Educational Research Quarterly, 11(2), 48-55.

Green, K. E., & Williams, J. E. (1989). Standardized test use by classroom teachers: Effects of training and grade level taught. Paper presented at the National Council on Measurement in Education annual conference, San Francisco.

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

Gullickson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. Journal of Educational Research, 79(2), 96-100.

Gullickson, A. R. (1986a). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23, 347-354.

Gullickson, A. R. (1986b). Perspective on educational measurement. Applied Psychology Measurement, 10, 109-132.

Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 4(1), 15-18.

Gullickson, A. R., & Hopkins, K. D. (1987). The context of educational measurement instruction for preservice teachers: Professor perspectives. Educational Measurement: Issues and Practice, 6, 12-16.

BEST COPY AVAILABLE

Haertel, E. (1986). Choosing and using classroom tests: Teachers' perspectives on assessment. Paper presented at the American Educational Research Association, San Francisco.

Hall, B. W., Carroll, D., & Comer, C. B. (1988). Test use among classroom teachers and its relationship to teaching level and teaching practices. Applied Measurement in Education, 1(2), 145-156.

Herman, J., & Dorr-Bremme, D. (1982). Assessing students: Teachers' routine practices and reasoning. Paper presented at the annual meeting of the American Educational Research Association, New York.

Hermanowicz, H. J. (1980). Toward a new model of teacher education. Paper presented at the Crossgate Seminar sponsored by the Pennsylvania Department of Education, Mechanicsburg, PA. (ERIC Document Reproduction Service No. ED 199215)

Hill, K. T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. Elementary School Journal, 85, 105-126.

Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). The effects of standardized testing. Boston: Kluwer-Nijoff.

Kuhs, T., Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1985). Differences among teachers in their use of curriculum embedded tests. The Elementary School Journal, 86, 141-153.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. Review of Educational Research, 58(1), 79-97.

Lambert, R. F. (1980-81). Teacher attitudes on testing: A multiple perspective. College Board Review, 118, 13-14, 29-30.

Lazar-Morrison, C., Polin, L., Moy, R., & Burry, J. (1980). A review of the literature on test use. CSE Report No. 144. Los Angeles, CA: Center for the Study of Evaluation.

Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. Teachers College Record, 91(3), 422-436.

Marso, R. N. (1969). The influence of test difficulty upon study efforts and achievement. American Educational Research Journal, 6(4), 621-632.

Marso, R. N. (1970a). Classroom testing procedures, test anxiety, and achievement. The Journal of Experimental Education, 38(3), 54-58.

Marso, R. N. (1970b). Test item arrangement, testing time, and performance. Journal of Educational Measurement, 7(2), 113-118.

Marso, R. N. (1985). Reliability and validity of essay and objective test items in a classroom setting. Educational Review, 9(1), 9-20.

Marso, R. N. (1986). Testing practices and test item preferences of classroom teachers. Paper presented at the Midwestern Educational Research Association, Chicago. (ERIC Document: ED 268 145)

Marso, R. N., & Pigge, F. L. (1987a). A state-wide assessment of the testing and evaluation needs and proficiencies of beginning teachers: Implication for staff development. A paper presented at the Association of Supervision and Curriculum Development, New Orleans. (ERIC Document: ED 283 833)

Marso, R. N., & Pigge, F. L. (1987b). Differences between self-perceived job expectations and job realities of beginning teachers. Journal of Teacher Education, 38(6), 53-56.

Marso, R. N., & Pigge, F. L. (1988a). An analysis of teacher-made tests: Testing practices, cognitive demands and item construction errors. A paper presented at the National Council on Measurement in Education, New Orleans. (ERIC Document: ED 298174)

Marso, R. N., & Pigge, F. L. (1988b). Ohio secondary teachers' testing needs and proficiencies: Assessments by teachers, 17(2), 2-9.

Marso, R. N., & Pigge, F. L. (1988c). Teacher-made tests and testing: Classroom resources, guidelines and practices. A paper presented at the Midwestern Educational Research Association, Chicago. (ERIC Document: ED 291 781)

Marso, R. N., & Pigge, F. L. (1988d). The availability in Ohio's schools of resources to support teacher-made testing: Surely we can do better! OASCD Dimensions, 9(5), 8-9.

Marso, R. N., & Pigge, F. L. (1989a). Elementary classroom teachers' testing needs and proficiencies: Multiple assessments and inservice training priorities. Educational Review, 13, 1-17.

Marso, R. N., & Pigge, F. L. (1989b). Standardized and competency testing programs: Classroom teachers' needs and proficiencies as viewed by principals and supervisors. Paper presented at the Midwestern Educational Research Association, Chicago. (ERIC Document: ED 304 473)

Marso, R. N., & Pigge, F. L. (1989c). The status of classroom teachers' test construction proficiencies: Assessment by teachers, principals, and supervisors validated by analyses of actual teacher-made tests. A paper presented at the National Council on Measurement in Education, San Francisco. (ERIC Document: ED 306 283)

Marso, R. N., & Pigge, F. L. (1990). Training, job titles, and responsibilities of directors of public school standardized testing programs. Paper presented at the National Council for Measurement in Education, Boston, MA.

Marso, R. N., & Pigge, F. L. (1991). An analysis of teacher-made tests: Testing practices, cognitive demands and item construction errors. Contemporary Educational Psychology, 16, 279-286.

Mehrens, W. A., & Lehmann, I. J. (1987). Using standardized tests in education (4th ed.). New York: Longman.

Monk, J. J., & Stallings, W. M. (1970). Effects of item order on test scores. Journal of Educational Research, 63, 463-465.

Newman, D. C., & Stallings, W. M. (1982). Teacher competency in classroom testing, measurement preparation, and classroom testing practices. Paper presented at the American Educational Research Association annual meeting, New York. (ERIC Document: ED 220 491)

Oescher, J., & Kirby, P. C. (1990). Assessing teacher-made tests in secondary math and science classrooms. Paper presented at the National Council on Measurement in Education, Boston.

Roeder, H. H. (1973). Teacher education curricula--your final grade is F. Journal of Educational Measurement, 10(2), 141-143.

Rogers, E. M. (1969). Examinations: Powerful agents for good or ill in teaching. American Journal of Physics, 37, 954-962.

Ruddell, R. B. (1985). Knowledge and attitudes toward testing: Field educators and legislators. The Reading Teacher, 38, 538-543.

Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening? Phi Delta Kappan, 62, 631-634.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38(3), 57-62.

Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. American Educational Research Journal, 14(2), 83-97.

Shulman, L. (1980). Test design: A view from practice. In E. Baker & E. Quellmalz (Eds.), Educational Testing and Evaluation. Beverly Hills, CA: Sage, 63-76.

Stetz, F. P., & Beck, M. D. (1981). Attitudes toward standardized tests: Students, teachers and measurement specialists. Measurement in Education, 12, 1-11.

Stiggins, R. J. (1985). Improving assessment where it means the most: In the classroom. Educational Leadership, 43, 69-74.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. Phi Delta Kappan, 69, 363-368.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practices, 5, 5-17.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. Journal of Educational Measurement, 26(3), 233-246.

Veenman, S. (1984). Perceived problems of beginning teachers. Review of Educational Research, 54(2), 143-178.

Yeh, J. P. (1981). Teachers and testing: A survey of test use. Washington, DC: U.S. Department of Health, Education and Welfare and National Institute of Education. (ERIC Document: ED 218 336)

PC:13A

Table 1

Educators' Attitude Toward K-12 Classroom Testing as Reflected in Their Beliefs, and in the Extent of Resources and Training Available to Support Classroom Testing

1. Just since 1990 have standards for classroom teachers' testing competence been available; whereas standards for standardized testing have existed since the middle of the century.

2. The educational and measurement communities generally believe that teacher-constructed tests have a greater impact upon instruction and pupil learning in classrooms than do other types of tests.

3. The measurement community perceive many teachers, as well as others in education, to have limited and inadequate classroom testing knowledge and skills.

4. Limited, if any, testing expertise is available in most K-12 school buildings to assist and support teachers' testing related responsibilities. Most undergraduate and graduate educational preparation programs do not require training in testing and measurement. Further, many college professors who instruct teacher candidates in tests and measurements, themselves, may have limited training or experiences in classroom testing.

5. Most K-12 educational administrators have little or no training in measurement, and limited emphasis is placed on the management of testing programs in the public schools.

6. Many K-12 classroom teachers have little or no formal training in tests and measurements. There are as many teacher preparation institutions requiring no formal measurement training or just requiring training as part of another course as there are institutions requiring a complete course in tests and measurement of their teacher candidates.

7. Most principals and teacher supervisors appear to neither value nor to encourage teacher use of technical testing skills such as use of item analysis, test specification tables, or test score statistical analysis procedures; teachers, themselves, also appear not to deem these skills to be essential to the success of their pupil testing efforts.

8. As many as 20 percent of the standardized testing directors of K-12 school districts have no more formal training in tests and measurements than the amount of training commonly expected of a beginning classroom teacher.

(table continues)

Table 1 (continued)

9. Even basic support services such as typing and duplication services are not consistently available in many schools to support teachers' testing responsibilities. Computerized support services such as scoring, item analysis, etc. are available in relatively even fewer schools.

10. Teachers report that inservice training related to classroom testing and measurement is rarely if ever available in their schools. Limited evidence suggests that neither teachers' perceptions of the quality of their testing proficiencies nor the observed quality of their self-constructed tests appear to improve with teachers' increased years of teaching experience.

11. Classroom teachers feel that teacher preservice training in tests and measurements is not designed to meet their needs, and educators generally do not participate in training related to testing unless it is required of them.

12. School principals and teacher supervisors rate beginning teachers' proficiencies in tests and measurements somewhat lower than they rate beginning teachers' proficiencies in other areas such as knowledge of subject content or classroom management related skills.

13. The measurement and education communities have conducted considerably less research on classroom teacher-devised testing compared to research of standardized testing and to many other aspects of education.

14. Limited research suggests that the increased availability of adequate support for school testing activities positively influences teachers' attitude toward testing.

Table 2

Classroom Teachers' Attitudes as Reflected in Their Beliefs About Testing and Their Classroom Testing Practices

1. Teachers believe that teacher-made tests generally have a positive impact upon the study and learning efforts of their pupils.

2. Teachers value and use assessment procedures that best meet their day-to-day instructional needs.

3. Teachers believe that teacher-made test assessments of their pupils should closely mirror the instruction being provided.

4. Teachers believe that for tests to be useful they must fit their instructional needs, must be of practical value, and test results must be immediately available.

5. Teachers believe, and indicate that school administrators and pupils also believe, that teacher-made tests should be scheduled on a relatively frequent basis to promote pupil learning.

6. Teachers believe that teacher-designed testing and the discussion of the results from these testing sessions are productive uses of classroom time.

7. Teachers believe that teacher-made tests are useful in diagnosing pupils' progress, making pupil grouping decisions, assigning pupil grades, and reporting the progress of pupils.

8. Teachers believe that self-constructed assessments generally better meet the instructional needs of their classes than do assessments derived from other sources such as workbooks and textbooks.

9. Teachers rely on teacher-made tests to a much greater extent than standardized tests and district or state competency tests for making decisions about individual pupils.

10. Teachers have a more positive attitude toward the value of teacher-made tests compared to standardized tests or state and school district pupil minimum competency tests.

11. Teachers believe that differing course content and pupil grade level variations require somewhat different assessment devices and practices.

12. Teachers believe that the results from formal tests should be supplemented by information from other sources such as observations and daily work when assigning grades or making other decisions about pupils.

(table continues)

Table 2 (continued)

13. Teachers believe that teacher-made test results aid teachers in justifying grades to pupils and the parents of pupils.

14. Teachers believe that daily classroom observations and teacher judgment are more reliable sources of information for making many classroom related decisions than are isolated test scores.

15. Teachers believe that test scores must be interpreted within the context of all other information available about a pupil.

16. Teachers believe that where pupil learning is displayed in overt behaviors less use should be made of paper and pencil type tests.

17. Most teachers place considerable value in and reliance on information about pupils gathered through informal observations, day to day communication, and daily work; teachers in the lower grades tend to rely more on these sources of information for making their classroom than upon formal tests while middle and upper grade teachers tend to rely more on formal tests than upon informally gathered information.

18. Teachers believe that essay tests and questions are less useful and less liked by pupils but that they result in greater pupil study efforts and usually measure at higher cognitive levels than do objective tests.

19. Teachers believe that matching, short-response, completion, and multiple-choice questions are the more useable, efficient, and useful types of questions in contrast to the essay or true-false question types.

20. Teachers believe that a variety of question types should be used in classroom tests in order to be fair to pupils and to better assess a variety of instructional objectives.

21. Teachers believe that teacher-made tests should contain questions that demand higher-order pupil thinking skills, and they tend to over estimate the cognitive demands of their tests.

22. Teachers believe that tests need to be administered fairly and efficiently and that teachers should monitor classes being tested to prevent pupil cheating.

23. Teachers believe that test results can be interpreted and conveyed to pupils adequately without the use of statistical analyses.

(table continues)

Table 2 (continued)

24. Teachers believe that testing and related assessment procedures, to be consistently used and useful in classrooms, must be efficient in time and energy demands of teachers as well as being supportive of on-going classroom instructional activities.

25. Teachers expend considerable in and out of class time and effort in testing and assessment activities. Teachers typically schedule formal tests once every two weeks or more often in most courses, construct on an average 54 formal tests each year, and construct most of their own test questions.

26. Teachers believe that testing, evaluation, and grading activities are among their more demanding and less pleasant classroom responsibilities.

27. Teachers commonly express concern about their pupil testing and evaluation responsibilities.

28. Teachers report spending little time editing or revising test questions prior to or following their use in tests.

Table 3

Classroom Teachers' Attitudes Toward Teacher-Made Tests and Testing as Reflected in Direct Analyses of Their Teacher-Made Tests

1. Teachers more frequently use short-answer, completion, and matching question types which commonly measure at the lower cognitive demand levels. Multiple-choice questions are also frequently used; true-false are used less often; and essay questions are used very infrequently.

2. Teachers express a valuing of items that measure at higher cognitive levels, but their tests measure predominantly at the knowledge cognitive functioning level.

3. Teachers appear to value the instructional aspects of teacher-testing but not the technical or skill aspects of test construction and use, for they display limited knowledge and proficiency in and seldom practice the technical aspects of testing (e.g., seldom use test specification tables, item analysis, item writing, and statistical analysis procedures, etc.).

4. Teachers appear to be unaware of, are unable to identify item writing flaws, or do not sufficiently value the use of common test writing guidelines, for analyses of teachers' tests consistently reveal very frequent violations of common test question and test format construction guidelines.

5. Teachers appear not to sufficiently value training in testing to improve testing skills, for neither inservice training, if provided, nor increased years of teaching experience appear to improve classroom teachers' testing knowledge and test construction skills.

BEST COPY AVAILABLE

Table 4

Classroom Teachers' Ratings of their Need for Selected Testing Competencies

	Testing Competencies or Skills	Mean* (N=313)	Rank
1.	Grading tests, papers, projects, homework, etc.	4.44	1
2.	Making tests reflect what is covered in text and class	4.35	2
3.	Calculating end of term grades from term work	4.29	3
4.	Identifying individual and class strengths and weaknesses	4.25	4
5.	Deciding importance of tests, papers, etc. in grading	4.23	5
6.	Determining what needs to be retaught after tests	4.20	6
7.	Constructing tests that represent true student progress	4.18	7
8.	Deriving information from tests to guide students	4.04	8
9.	Identifying good and poor questions for future tests	4.03	9.5
10.	Use of observations (visual) to assess and guide learning	4.03	9.5
11.	Writing questions in harmony with school and class goals	4.01	11
12.	Interpreting test scores and student progress	4.00	12
13.	Use of tests and grades to positively influence learning	3.99	13
14.	Setting up readable, scorable, and attractive tests	3.94	14
15.	Stating objectives sufficiently clear to suggest test items	3.88	15
16.	Writing test questions that demand higher thinking processes	3.81	16
17.	Selecting good test questions from teacher manuals	3.54	17.5
18.	Writing good matching questions	3.54	17.5
19.	Writing good completion questions	3.53	19
20.	Writing good multiple-choice questions	3.33	20
21.	Writing good true-false questions	3.31	21.5
22.	Use of less formal assessments: checklists, ratings, etc.	3.31	21.5
25.	Use of sociometric, guess who, and related techniques	2.71	25
23.	Scoring essay questions	3.24	23
24.	Writing good essay questions	3.20	24
26.	Calculation of means, standard deviations, reliability, etc.	2.49	26

*Means were derived from a 5-point Likert scale where 5 = high.